Two-Point Quasifractional Approximant in Physics. Truncation Error.

Pablo Martín,

Depto. de Física, Universidad Simón Bolívar, Apartado 89000, Caracas, Venezuela.

George A. Baker, Jr.

Theoretical Division, Los Alamos National Laboratory,

University of California, Los Alamos, N.M. 87545, U.S.A.

Abstract: The quasifractional approximation method is developed in a systematic manner. This method uses simultaneously the power series, and at a second point, the asymptotic expansion. The usual form of the approximants is two or more rational fractions, in terms of a suitable variable, combined with auxiliary non-fractional functions. Coincidence in the singularities in the region of interest is pursued. Equal denominators in the rational fractions is required so that the solution of only linear algebraic equations is needed to determine the parameters of the approximant. An upper bound is obtained for the truncation error for a certain class of functions, which contains most of the functions for which this method has been applied so far. It is shown that quasifractional approximants can be derived as a mixed German and Latin polynomial problem in the context of Hermite-Padé approximation theory.

Suggested PACS numbers: 02.60.+y, 02.30.+g.

I. Introduction and Summary

Padé approximation has been successfully applied in several areas of physics. The theory and the method of their computation are well established.[1−3] We want, however, to point out some limitations that are important and to show why there still remain some problems in physics for which they are not adequately well suited. Physicists are use to schemes of perturbation where the leading term is the main contribution to the problem and the new terms are just improvements that give better accuracy in the determination of the function or parameters under study. This situation occurs, for instance, in the WKB method, the Born approximation, perhaps in quantum field theory, *etc.* On the other hand, the Padé approximants are based on the power series expansion at a point, or at several points, as in the multipoint method. One of these points could, of course, be the point at infinity, as in the case which is sometimes called the Padé-Laurent method. One difficulty with the power series expansion is that frequently the first term is not a good approximation, and furthermore, it very often happens that each new term is more important than the preceding one, until a given power is reached. After that, each term is descending in value and less important. It may, of course, also happen that the terms just keep increasing. For these cases, resummation, even beyond the radius of convergence by the method of Padé approximants, has often been very successful. Never-the-less it is worth while to search for additional methods which come closer the basic principal of approximation theory, which is to build into the structure of the approximation every thing that you *really know for sure* that you conveniently can. Interestingly, there is a peculiar expansion, or quasi-series that has the above mentioned characteristic of the usual scheme of perturbation in physics–namely, the well known asymptotic expansion of special functions. A peculiarity of this expansion is that the Poincaré extension of the convergence domains gives an appropriate criterion for the truncation error. Asymptotic expansions are used so extensively in physics that the references include practically all the graduate text books in physics.[4,5] It seems therefore appropriate to define a new system of fractional approximants based on both the convergent power series and the asymptotic expansion about a different point. In its implementation it will be distinct from the two point Padé approximant, but will, as described in the fourth section of this paper, represent the one-point Latinization of the German Hermite-Padé two-point approximants (vector valued rational interpolants). From the beginning we can say that this task is difficult and full of obstacles. The pattern of the asymptotic expansion is not so simple as that of the power series. Here we have to consider not only an ensemble of increasing integer powers, but furthermore additional factors in this ensemble that usually are fractional powers, exponential or trigonometric functions, but could in principle be practically any function. Furthermore, we do not have here a circle of convergence but instead, the form of the asymptotic expansion will depend on the sector in the complex plane, and the occurrence of Stokes phenomena must be considered. In several previous papers we have shown practical ways of treating this problem for several particular cases.[6−11] This paper discusses the general procedure to be employed to obtain these approximants, the "two-point quasifractional approximants."

By looking at the general problem, it has been possible to find a way to bound the truncation error without resorting to the procedure of computing the difference between the approximant and the exact function or parameter. The proof depends on certain assumed properties which must be verified to use the results for a particular case. We have analyzed the truncation error for the particular case of the first order approximant to the Bessel function. These results are not the best possible, but represent a beginning. We want to emphasis that in the one-point Padé approximant case the determination of the error of approximation is a problem of considerable difficulty and long standing.

The philosophy of this paper is different from our previous ones. Here the ideas of the method will be emphasized. No new particular computation is included. On the other hand a formula for the truncation error has been found that will be useful when the exact function or parameter (eigenvalue) can not be determined through direct computation.

The material of the paper has been arranged as follows. In Section II we discuss the general procedure required to obtain the two-point quasifractional approximants. The analysis of the truncation error is done in Section III, including some results. Section IV discusses the relationship between the two-point quasifractional approximants and the general Hermite-Padé approximants of both German and Latin types. Specifically they are a variant of the vector-valued rational interpolants.

As in the two-point Padé method, we use information coming from two points, but here we utilize the power series at one point and the asymptotic expansion at the other.

II. Theoretical Treatment

In order to simplify the analysis, we will consider a function determined by a power series expansion at zero and having an asymptotic expansion at infinity. This specialization is not a limitation, since the results can easily be extended to at function for which the origin and infinity could equally well be replaced by any two points. Thus we assume that $f(z)$ is a function such that,

$$f(z) = z^r \sum_{i=0}^{N} a_i z^i, \quad |z| \ll 1, \tag{2.1}$$

$$f(z) = A(z) \left[B_1(z) \sum_{j=0}^{M} b_j z^{-j} + B_2(z) \sum_{j=0}^{L} c_j z^{-j} \right]_{\alpha < \arg z < \beta}, \quad |z| \gg 1. \tag{2.2}$$

In this analysis we are giving a determined form for the asymptotic expansion that is appropriate for a lot of asymptotic expansions. More complicated forms can be treated using an analysis similar to that describe here. The functions $A(z)$, $B_1(z)$ and $B_2(z)$ are usually fractional powers, exponential or trigonometric functions, but any other functions could be considered. We are denoting by r a given number, not an index, and it usually the number unity or a simple rational number like $\frac{1}{2}$, $\frac{1}{4}$, $\frac{3}{4}$, *etc.* Due to Stokes phenomena, the asymptotic expansion is normally limited to a circular sector of the complex z-plane.

The function $f(z)$ could be given by the perturbation expansion of an eigenvalue problem where only a few terms of the expansion are known. When the complete power series or asymptotic expansion is known the numbers N, M and L are infinite. We will assume that the functions $A(z)$, $B_1(z)$ and $B_2(z)$ are bounded for large values of z. In the cases where this condition is not satisfied, the infinite part of the factor will be included in the function $f(z)$. For instance if $A(z) = e^z$, the analysis will be done for the function

$e^{-z}f(z)$ instead of $f(z)$.

The problem now of finding a two-point quasifractional approximation to this function $f(z)$, valid for any real positive value of z will be considered in several steps. The usual situation is that the functions $A(z)$, $B_1(z)$ and $B_2(z)$ are not bounded in the open interval, from zero to infinity. In the cases that one of these functions has an negative power of z as an over all factor, that function will blow up for $z = 0$. As this point is in our region of interest, it is better to extend the region of validity of our approximant to include the neighborhood of the origin, to insure good behavior at $z = 0$.

In regard to the preceding paragraph, the first step is to replace the functions $A(z)$, $B_1(z)$ and $B_2(z)$ by appropriate functions, $\tilde{A}(z)$, $\tilde{B}_1(z)$ and $\tilde{B}_2(z)$. These new functions are to be bounded in our region of interest and furthermore they should have an appropriate asymptotic behavior. This statement usually means that the leading terms of the corresponding functions are coincident. To be clearer, the case of a fractional power is analyzed in detail. Let us assume that $A(z)$ is given by

$$A(z) = 1/z^s. \tag{2.3}$$

Our new function $\tilde{A}(z)$ could be

$$\tilde{A}(z) = \frac{1}{(1+z)^s}. \tag{2.4}$$

Here we are introducing a branch point in the approximant at $z = -1$, which is not in the function $f(z)$, however this point is outside of our region of interest. In popular terms, this is the price we have to pay in order to keep the approximant bounded. We are in some sense free to choose the point for the new singularity which must be outside of the neighborhood of our region of interest and not very close to that region.

The golden rule in our analysis is that the function and the quasifractional approximant must have the same singularities in the region of interest, but we can select points outside of the region of interest to put all the new singularities.

The second step is the selection of the number of fractions of our approximants and their denominators. In the case of our example, we have to use two fractions, but more complicated situations will lead to a larger number of fractions. The denominator of both fractions can be chosen equal or different. If both denominators are predetermined, we do not have to be careful. However when we allow free parameters in the denominators, we have to choose the part with free parameters equal in both fractions in order that the calculation of the parameters can be accomplished by the use of *linear* algebraic equations only. It is a good technique to predetermine the denominators is such a way as to avoid the occurrence of "defects" (*i.e.*, a close pole-zero pair) which are known to occur frequently in Padé analysis. In our past experience, it has proven to be a good procedure to predetermine the denominator and to put all the new singularities at the point $z = -1$, or on the circle $|z| = 1$. The first case leads to denominators of the form $(1+z)^m$ and the second one to the forms $(1+z^l)^k$, such that $kl = m$. However, the selection of the denominator depends on the knowledge we have of the function $f(z)$. If we know that function has a given pole in the region of interest, that pole must be included in the denominator as a factor and we are free choose the other factors. Even if the poles of $f(z)$ are outside the region of interest, it is convenient to include it in the approximant, if we know where they are. In our procedure, we should try to introduce all that sort of information through an appropriate definition of the denominator.

The third step in our procedure is to introduce suitable auxiliary functions, $\tilde{A}(z)$, $\tilde{B}_1(z)$ and $\tilde{B}_2(z)$ in such a way that the determination of the parameters of the approximant can be done without contradictions, which requires similar behavior for the approximant and the function around both zero and infinity. In our specific example the approximant can be defined as,

$$\hat{f}_1(z) = \frac{z^r}{(1+z)^r}\tilde{A}(z)\left[\tilde{B}_1(z)\frac{\sum_{i=0}^{n} p_i z^i}{(1+z)^n} + \tilde{B}_2(z)\frac{\sum_{i=0}^{n} P_j z^j}{(1+z)^n}\right]. \tag{2.5}$$

Clearly the additional factor $z^r/(1+z)^r$ is introduced in such a way that the approximant has the branch point at $z = 0$, but no new undesirable singularity is introduce at z infinite.

The parameters to be defined here are the p's and the P's. To do that we have to select an appropriate number of terms from the expansions in Eq. (2.1) and (2.2). Usually we choose an equal number of terms from $(\sum b_j z^{-j})$ and from $(\sum c_k z^{-k})$. (In some cases we will use one more term from one of the expansions however.) If we choose m terms from the power series, and l and t terms respectively from each of the asymptotic expansions, we determine the parameters for the approximant through the equations resulting from the relations,

$$z^{-r} f(z) = z^{-r} \hat{f}(z) + O(z^m), \tag{2.6}$$

$$f(z) - \frac{(1+z)^r A(z)}{z^r \tilde{A}(z)} \hat{f}(z) = A(z)[B_1(z) O_1(z^{-l-1}) + B_2(z) O_2(z^{-t-1})], \tag{2.7}$$

where the usual notation for functions of order z^m, *etc.* is used. Now the relation between n, m, l and t is

$$2(n+1) = m + l + t + 3. \tag{2.8}$$

In most cases l and t will be taken equal, but this is not required. We assume that the power series starts with z^0 and also that the asymptotic expansions do as well. In this way we will obtain the same number of equations as unknowns. In other cases, the modifications are easily done.

In case the denominator is undetermined, we can define the quasifractional approximant as

$$\hat{f}(z) = \frac{z^r}{(1+z)^r} \tilde{A}(z) \left[\tilde{B}_1(z) \frac{\sum_{i=0}^{n} \bar{p}_i z^i}{\sum_{k=0}^{n} q_k z^k} + \tilde{B}_2(z) \frac{\sum_{i=0}^{n} \bar{P}_i z^i}{\sum_{k=0}^{n} q_k z^k} \right], \tag{2.9}$$

where we have now $3n$ unknowns and the relation between the n, m, l and t will be

$$3n + 2 = m + l + t + 3. \tag{2.10}$$

A specific example of the preceding analysis is the Bessel function $J_\nu(x)$ of integer or fractional order. Our preceding analysis however is incomplete because sometimes the variable in the asymptotic expansion is different from that in the power series. This situation leads to an additional step in our procedure which consists in finding a suitable expansion variable for the quasifractional approximant in such a way that its power series and asymptotic expansion has the form of the exact function. The way to do that in a general way is not clear as yet, however for some specific examples the procedure can be illustrated. In the case of elliptic functions, the constraints involved in finding suitable variables lead to a differential equation of the Riccati type whose solution is found to be $\mathrm{tg}(z/2)$.

In the case of the Airy function, a first analysis shows that a suitable variable will be ζ^2, where $\zeta = \frac{2}{3}x^{3/2}$. Clearly ζ can not be a suitable variable since ζ has an undesirable branch point at $x = 0$, however ζ^2 does not have that branch point. Therefore the choice of ζ^2 as an independent variable is preferable. In this case, we need the power $x^{3/2}$, so we use $\sqrt{1+\zeta^2}$. In this way the undesirable branch points will be off the positive real axis and not very close to it. Therefore the form of the approximant on the positive real axis will be,

$$\widehat{Ai}(x) = \left[\frac{\sum_{i=0}^{n} p_i \zeta^{2i}}{\sum_{k=0}^{n} q_k \zeta^{2k}} + \frac{x}{(\lambda^2+\zeta^2)^{2/3}} \frac{\sum_{i=0}^{n} P_i \zeta^{2i}}{\sum_{k=0}^{n} q_k \zeta^{2k}}\right] \frac{\exp[-\sqrt{\lambda^2+\zeta^2}]}{(\lambda^2+\zeta^2)^{1/12}}, \tag{2.11}$$

$$\zeta = \tfrac{2}{3}x^{3/2}, \quad x > 0. \tag{2.12}$$

Here we are leaving a real free parameter, λ, which can be chosen arbitrarily. For instance, we can choose $\lambda = 1$. The only rule about λ is that it should not be very small in order to avoid the undesirable branch point getting to close to $x = 0$, a point in our region of interest.

Our notation in Eq. (2.11) is to denote the approximant with the same letter as that for the corresponding function, but with a circumflex above it.

The preceding form for the approximant to the Airy function has adequate behavior at zero and at infinity, when this point is approached along the positive real axis. Furthermore, its power series expansion has only powers of the type z^{3k} or z^{3k+1} as does its corresponding function. Therefore, it is an approximant with good efficiency. This last word is used in the sense that there is a correspondence between the number of parameters to be determined and the number of terms to be used of the power series and the asymptotic expansion. In the preceding equation, we have $(3n+2)$ parameters to be determined and this number should be equal to the sum of the number of terms to be used from the power series and from the asymptotic expansion. The parameters will be determined from the linear algebraic equations obtained after the multiplication by $\sum_{k=0}^{n} q_k \zeta^{2k}$ and the use of the accuracy-through-order principle on the difference between the exact function and the approximant.

Finally it is important to point out that if we want to avoid the defect problem, that is, an extraneous pole and a nearby zero, then we can prescribe the denominator in a suitable way. For instance, we can write it as

$$\sum_{k=0}^{n} q_k \zeta^{2k} = (1+\zeta^2)^n, \tag{2.13}$$

and in this case we have only $(2n+2)$ parameters to be determined.

The case of the negative axis does not present any special problems. The suitable variable will again be ζ^2, but now ζ will be defined as

$$\zeta = \tfrac{2}{3}(-x)^{3/2}, \quad x \leq 0, \tag{2.14}$$

and the form of the approximant will be

$$\widehat{Ai}(x) = \left[\frac{\sum_{i=0}^{n} \bar{p}_i \zeta^{2i}}{\sum_{k=0}^{n} q_k \zeta^{2k}} \cos\zeta + \frac{x}{(\lambda^2+\zeta^2)^{2/3}} \frac{\sum_{i=0}^{n} \bar{P}_i \zeta^{2i}}{\sum_{k=0}^{n} q_k \zeta^{2k}} \frac{\sin\zeta}{\zeta} \right] (\lambda^2+\zeta^2)^{-1/12}. \tag{2.15}$$

Here we can make the same considerations as before in the sense that λ should not be very small and that we can predetermine the denominator as in Eq. (2.13).

The preceding example illustrates some of the difficulties with the this new method of two-point quasifractional approximation. However, in some problems the determination of a suitable variable can be more complicated, as in the case with a non-linear equation. One case where we can see the details of the problem is the case of elliptic functions where we have shown that the search for a suitable variable leads to the need to solve an auxiliary differential equation of the Riccati type.

Another example that is useful to illustrate our method is the case of Bessel functions of fractional order. We know that these functions have a branch point of order ν at the origin times a power series in the square of the argument, and that they have an asymptotic expansion at infinity in terms of one over the square of the variable, multiplied by a trigonometric function. Our approximant will be of the form

$$\hat{J}_{\nu,(1)}(x) = \frac{x^{\nu}}{(1+x)^{\nu+\frac{1}{2}}} \left[\frac{\sum_{i=0}^{n} P_i x^i}{\sum_{k=0}^{n} q_k x^k} \cos x + \frac{\sum_{i=0}^{n} p_i x^i}{\sum_{k=0}^{n} q_k x^k} \sin x \right]. \tag{2.16}$$

This form will give the desired conditions at the origin, and at infinity. We could also replace the denominator for a given polynomial of degree $2n$ with singularities in the left-hand plane or off the real axis. In this case we do not have any problem with "defects." Another possible form for the approximant is

$$\hat{J}_{\nu,(2)}(x) = \frac{x^{\nu-1}}{(1+x)^{\nu-\frac{1}{2}}} \left[\frac{\sum_{i=1}^{n} \bar{P}_i x^i}{\sum_{k=0}^{n} \bar{q}_k x^k} \cos x + \frac{\sum_{i=0}^{n} \bar{p}_i x^i}{\sum_{k=0}^{n} \bar{q}_k x^k} \sin x \right] \tag{2.17}$$

however, both forms have the problem that although they are of the correct structure at $x = \infty$, they have a power series expansion of the form $x^{\nu} \sum a_n x^n$, while the Bessel series is of the form $x^{\nu} \sum a_m x^{2m}$, so that all the odd powers of the approximant matched at the origin should be zero. A more efficient form in the sense already discussed would necessarily be more complex. One possible form would be

$$\hat{J}_{\nu,(3)}(x) = \frac{x^{\nu}}{(1+x^2)^{\frac{\nu}{2}+\frac{1}{4}}} \left[\frac{\sum_{i=0}^{n} \bar{\bar{P}}_i x^{2i}}{\sum_{l=0}^{n} \bar{\bar{q}}_k x^{2k}} \cos \chi + \frac{\sum_{i=0}^{n} \bar{\bar{p}}_i x^{2i}}{\sum_{l=0}^{n} \bar{\bar{q}}_k x^{2k}} \frac{\sin \chi}{x} \right], \tag{2.18}$$

where, when we choose

$$\chi = x \left[1 - \frac{(\frac{1}{2}\nu + \frac{1}{4})\pi}{\sqrt{(1+x^2)}} \right], \tag{2.19}$$

we obtain the form of Hankel's asymptotic expansion at $x = \infty$ and a power series of the correct form at $x = 0$. This latter form has not yet been studied carefully numerically.

Having illustrated how to determine the form of the two-point quasifractional approximant, we will consider the truncation error in the next section.

III. Truncation Error

In order to find a procedure to evaluate the truncation error, we will refer to the general form of the function [eq. (2.1) and (2.2)] for which we have found an approximant given by eq.(2.9). We will restrict our analysis now to the more usual case where $l = t$. The extension to the case where t differs from l by a number of order unity is straightforward. From the conditions imposed on the parameters, we know that

$$z^{-r}[f(z) - \hat{f}(z)] = O(z^{m+1}),\ 0 \leq z < \infty. \tag{3.1}$$

We will denote by $\phi(z)$ and $\hat{\phi}(z)$ the functions,

$$\phi(z) = z^{-r} f(z),\ \hat{\phi}(z) = z^{-r} \hat{f}(z). \tag{3.2}$$

Now from the properties of the remainder of the Taylor series, we know that

$$\phi(z) - \hat{\phi}(z) = \frac{\phi^{(m+1)}(\theta z)}{(m+1)!}(\theta z)^{m+1} - \frac{\hat{\phi}^{(m+1)}(\tilde{\theta} z)}{(m+1)!}(\tilde{\theta} z)^{m+1}, \tag{3.3}$$

where θ and $\tilde{\theta}$ are numbers in the interval [0,1], *i.e.* $0 \leq \theta \leq 1;\ 0 \leq \tilde{\theta} \leq 1$.

The first restriction that we are going to impose on our functions is that the difference $\phi^{(m+1)}(\theta z) - \hat{\phi}^{(m+1)}(\tilde{\theta} z)$ is bounded. Thus we will assume,

$$\frac{|\phi^{(m+1)}(\theta z) - \hat{\phi}^{(m+1)}(\tilde{\theta} z)|}{(m+1)!} \leq M(m),\ 0 \leq z \leq \infty \tag{3.4}$$

where M is a positive number that depends on m, but is independent of z.

From our preceding equations we have the first restriction,

$$|f(z) - \hat{f}(z)| \leq M(m) z^{m+r+1},\ 0 \leq z \leq \infty. \tag{3.5}$$

Looking now at the asymptotic expansions, we have here a Poincaré type convergence. On the other hand, the conditions imposed on the parameters lead to

$$\sum_{j=1}^{\infty} b_j z^{-j} - \frac{1}{A(z)B_1(z)} \frac{z^r}{(1+z)^r} \tilde{A}(z)\tilde{B}_1(z) \frac{\sum_{i=0}^{n} \bar{p}_i z^i}{\sum_{k=0}^{n} q_k z^k} = O(z^{-(l+1)}). \tag{3.6}$$

We will assume as before that our functions are bounded. Thus we can find numbers $M_1(l)$ and $M_2(l)$ which are functions only of l and not of z such that the second restriction,

$$\left| \sum_{j=l+1}^{\infty} b_j z^{-j} \right| \leq M_1(l) z^{-(l+1)},\ 0 \leq z \leq \infty, \tag{3.7}$$

where the sum on j means the complete sum minus a polynomial to maintain a consistent definition for asymptotic series, and the third restriction,

$$\left| \frac{1}{A(z)B_1(z)} \frac{z^r}{(1+z)^r} \tilde{A}(z)\tilde{B}_1(z) \frac{\sum_{i=0}^{n} p_i z^i}{\sum_{k=0}^{n} q_k z^k} - \sum_{j=0}^{l} b_j z^{-j} \right| \leq M_2(l) z^{-(l+1)},$$
$$0 \leq z \leq \infty, \tag{3.8}$$

hold. The corresponding restrictions for the second series will serve to determine the numbers such that the fourth restriction,

$$\left| \sum_{j=t+1}^{\infty} c_j z^{-j} \right| \leq M_3(t) z^{-(t+1)},\ 0 \leq z \leq \infty, \tag{3.9}$$

and fifth restriction,

$$\left| \frac{1}{A(z)B_2(z)} \frac{z^r}{(1+z)^r} \tilde{A}(z)\tilde{B}_2(z) \frac{\sum_{j=0}^{n} \bar{P}_j z^j}{\sum_{k=0}^{n} q_k z^k} - \sum_{j=0}^{t} c_j z^{-j} \right| \leq M_4(t) z^{-(t+1)},$$
$$0 \leq z \leq \infty, \tag{3.10}$$

also hold.

We will determine a sixth restriction from the usual form of the asymptotic expansions, where $A(z)$ is a fractional power, and $B_1(z)$ and $B_2(z)$ are trigonometrical with decreasing exponential powers. Therefore we will also restrict our analysis with the following conditions,

$$|A(z)B_1(z)| \leq M_5 z^s,\ |A(z)B_2(z)| \leq M_5 z^s,\ 0 \leq z \leq \infty, \tag{3.11}$$

where usually s is a number less that unity. Considering all our preceding restrictive assumptions, we have that,

$$|f(z) - \hat{f}(z)| \leq M_5[M_1(l) + M_2(l) + M_3(l) + M_4(l)]z^{-(l+1)+s}, \tag{3.12}$$

which can be written as

$$|f(z) - \hat{f}(z)| = \tilde{M}(l)z^{-(l+1)+s},\ 0 \leq z \leq \infty, \tag{3.13}$$

$$\tilde{M}(l) = M_5[M_1(l) + M_2(l) + M_3(l) + M_4(l)]. \tag{3.14}$$

Now the inequalities (3.5) and (3.13) lead to a limit for the truncation error, since the largest error will be at the point where the right-hand-sides of both equations are equal. If we denote that point as z_0, we have

$$M(m)z_0^{m+r+1} = \tilde{M}(l)z_0^{-(l+1)+s}, \tag{3.15}$$

$$z_0 = \left[\frac{\tilde{M}(l)}{M(m)}\right]^{(m+l+2+r-s)^{-1}}. \tag{3.16}$$

If we denote the largest error by ϵ, we can bound it with,

$$\epsilon \leq [M(m)]^{\frac{l+1-s}{m+l+2+r-s}}[\tilde{M}(l)]^{\frac{m+1+r}{m+l+2+r-s}}. \tag{3.17}$$

Therefore we have gotten a limit for the largest possible error which can occur for the two-point quasifractional method, when all the six preceding restrictive assumptions are verified. An important limitation in our analysis is when $q(x)$ has a zero. It could reflect the presence of a defect, *i.e.* a nearby pole and zero. In this case the restriction can not be satisfied. If we determine the denominator in such a way that the zeros of $q(x)$ are out of the region of interest, we can avoid this problem.

In order to understand the above methods of analysis for the truncation error, we will consider the particular case of the fractional order Bessel function for the first approximant form discussed herein at eq. (2.15). We will consider the first order approximant as in a previous paper.[8] Using our previous notation, we have

$$
\begin{gathered}
r=\nu,\ s=-\tfrac{1}{2},\ \alpha=(2\nu+1)\frac{\pi}{4},\ n=1,\ l=0,\ m=2, \\
A(z)=\frac{1}{\sqrt{x}},\ \tilde{A}(z)=\frac{1}{\sqrt{1+x}}, \\
B_1(z)=\tilde{B}_1(z)=\cos x,\ B_2(z)=\tilde{B}_2(z)=\sin x.
\end{gathered}
\tag{3.18}
$$

Since $m=2$, we have to consider the third derivative in our analysis of the power series. To determine M, we have to calculate the upper bound for

$$
\frac{d^3}{dx^3}[x^{-\nu}J_\nu(x)],
\tag{3.19}
$$

for $x=\theta z_0$, and

$$
\frac{d^3}{dx^3}[x^{-\nu}\hat{J}_\nu(x)],
\tag{3.20}
$$

for $x=\tilde{\theta}z_0$, where z_0 is the point at which the maximum error occurs. The important remark is that the series for these functions are alternating series. The terms of the series are of the form $a_k z^k$, where the factors a_k are decreasing coefficients. We are interested in knowing when each new term of the series is smaller than the preceding one. The series for $J_\nu(x)$ is well known. Thus we can proceed to compare the coefficients in the third derivative. The power series for $J_\nu(z)$ is in powers of z^2. The ratio of successive terms is

$$
\begin{aligned}
\frac{a_{k+1}z^{2(k+1)}}{a_k z^{2k}} &= \frac{(2k+2)(2k+1)(2k)(-4)^{-k-1}z^{2k-1}k!\Gamma(\nu+k+1)}{2k(2k-1)(2k-2)(-4)^{-k}z^{2k-3}(k+1)!\Gamma(\nu+k+2)} \\
&= -\frac{(2k+1)z^2}{4(2k-1)(k-1)(\nu+k+1)}.
\end{aligned}
\tag{3.21}
$$

For our analysis, $k=2$, and the worst situation is for $\nu=-1$, among the cases that we will be considering. Thus, when $z \lesssim 2.2$, the first term will be a bound for all the rest of the series. So, if $\theta z_0 \lesssim 2.2$, we can keep only the first term, $a_k\theta^k z^k$. We will show later that this condition is well verified in our example. Thus we can keep this bounded value. Since $0<\theta<1$, we can also obtain a bound by the use of the simple term $a_k z^k$. A similar analysis can also be applied to the power series of (3.20) for $x = \tilde{\theta} z_0$, but now the power series has all the powers and not just the even ones. When we combine both series together we find that

$$x^\nu |J_\nu(x) - \hat{J}_\nu(x)| \le x^\nu \left[\mathcal{C}_1 z_0^4 + \mathcal{C}_2 z_0^3\right], \quad x \le z_0, \quad \theta z_0, \; \tilde{\theta} z_0 \lesssim 2.2. \tag{3.22}$$

where

$$\mathcal{C}_1 = \frac{1}{2^{\nu+5}\Gamma(\nu+3)}, \tag{3.23}$$

$$\begin{aligned}\mathcal{C}_2 = \Bigg| \frac{P_0}{q_0}\Bigg[&\binom{-\nu-\frac{1}{2}}{3} + \binom{-\nu-\frac{1}{2}}{2}\left(\frac{P_1}{P_0} - \frac{q_1}{q_0}\right) \\ &- \binom{-\nu-\frac{1}{2}}{1}\left(\frac{1}{2} + \frac{q_1 P_1}{q_0 P_0} - \frac{q_1^2}{q_0^2}\right) + \left(\frac{P_1}{P_0} - \frac{q_1}{q_0}\right)\left(\frac{q_1^2}{q_0^2} - \frac{1}{2}\right)\Bigg] \\ &+ \frac{p_0}{q_0}\left[\binom{-\nu-\frac{1}{2}}{2} + \binom{-\nu-\frac{1}{2}}{1}\left(\frac{p_1}{p_0} - \frac{q_1}{q_0}\right) - \left(\frac{1}{6} + \frac{p_1 q_1}{p_0 q_0} - \frac{q_1^2}{q_0^2}\right)\right]\Bigg|,\end{aligned}$$

where $p_0, \; p_1, \; P_0, \; P_1, \; q_0$ and q_1 are given in eq. (4) of reference (8). The number M is defined as

$$M = \mathcal{C}_1 + \mathcal{C}_2. \tag{3.24}$$

Now we have to consider how the result depends on whether z_0 is smaller or larger than unity, since in one case z_0^3 will be larger and in the other case smaller than z_0^4. Thus,

$$x^\nu |J_\nu - \hat{J}_\nu| \le \begin{cases} M z_0^{\nu+3} & \text{for} \quad z_0 \le 1 \\ M z_0^{\nu+4} & \text{for} \quad z_0 > 1 \end{cases} \tag{3.25}$$

Now we can proceed to the analysis of $\tilde{M}$, which is simpler. A similar argument based on the known properties of Hankel's asymptotic expansion can be given to justify using the truncation of the series to obtain the bounding values.

The other bounding parameters are determined by use of the asymptotic expansions. Since in this case we use only the zeroth order terms ($l = 0$), we have only to expand to order $(1/x)$. Thus,

$$
\begin{aligned}
M_1 &= \sqrt{\frac{2}{\pi}}\frac{|(4\nu^2-1)\cos\alpha|}{8},\\
M_2 &= \sqrt{\frac{2}{\pi}}|\cos\alpha|\left|\frac{P_0}{P_1}-\frac{q_0}{q_1}-(\nu+\tfrac{1}{2})\right|,\\
M_3 &= \sqrt{\frac{2}{\pi}}\frac{|(4\nu^2-1)\sin\alpha|}{8},\\
M_4 &= \sqrt{\frac{2}{\pi}}|\sin\alpha|\left|\frac{p_0}{p_1}-\frac{q_0}{q_1}-(\nu+\tfrac{1}{2})\right|,\\
M_5 &= 1,
\end{aligned}
\tag{3.26}
$$

and so,

$$
\begin{aligned}
\tilde{M} =& \sqrt{\frac{2}{\pi}}\Bigg\{\frac{|4\nu^2-1|}{8}(|\cos\alpha|+|\sin\alpha|)+|\cos\alpha|\left|\frac{P_0}{P_1}-\frac{q_0}{q_1}-(\nu+\tfrac{1}{2})\right|\\
&+|\sin\alpha|\left|\frac{p_0}{p_1}-\frac{q_0}{q_1}-(\nu+\tfrac{1}{2})\right|\Bigg\}.
\end{aligned}
\tag{3.27}
$$

The error bound ϵ has the value,

$$
\epsilon(\nu) \le \begin{cases} M(\nu)^{\frac{3}{9+2\nu}}\tilde{M}(\nu)^{\frac{6+2\nu}{9+2\nu}} & \text{for} \quad z_0 \le 1 \\ M(\nu)^{\frac{3}{11+2\nu}}\tilde{M}(\nu)^{\frac{8+3\nu}{11+2\nu}} & \text{for} \quad z_0 > 1 \end{cases}. \tag{3.28}
$$

In the case when $\nu < 0$, the calculation of the error was done for the function $x^{-\nu}J_\nu(x)$ instead of $J_\nu(x)$ (See ref. 10). Thus in this case the largest error $\tilde{\epsilon}(\nu)$ will be given by

$$
\tilde{\epsilon}(\nu) \le \begin{cases} M(\nu)^{\frac{3+2\nu}{9+2\nu}}\tilde{M}(\nu)^{\frac{6}{9+2\nu}} & \text{for} \quad z_0 \le 1 \\ M(\nu)^{\frac{3+2\nu}{11+2\nu}}\tilde{M}(\nu)^{\frac{8}{11+2\nu}} & \text{for} \quad z_0 > 1 \end{cases}. \tag{3.29}
$$

In Fig. 1 we can compare our maximum truncation error $\epsilon(\nu)$ with the maximum error found by direct calculation of the approximant and the exact Bessel function. We have divided the truncation error by ten in order to get a clearer figure. As we can see from the figure, the patterns of both curves are similar, however our truncation error is about ten times larger than the real error. Probably a way to get a better result is to get better values for M and $\tilde{M}$. In any case, since the pattern is always clearly larger than the real error, these results are in agreement with the theory as previously developed.

It is interesting to point out that in the figure for most of the values of ν, the point of maximum error is $1 < z_0 \lesssim 1.7 < 2.2$. We have illustrated that by using a full line when $z_0 > 1$. Near the points -0.5 and 0.5, $z_0 \leq 1$. This part is shown in the figure by the broken lines. The maximum errors obtained by the difference of the approximant and the exact function [10] are represented by a point-dash line. For negative ν the errors are those of the functions $x^{-\nu} J_\nu(x)$ and $x^{-\nu} \hat{J}_\nu(x)$, in order to avoid problems with the infinite value of $J_\nu(x)$ at $x = 0$.

IV. Relation to Hermite-Padé Approximants

In this section we consider the relationship between the various kinds of quasifractional approximants and the general family of Hermite-Padé[12,13] approximants. Suppose we are given a system of m functions $(F_0, \ldots, F_1)$ which possess power series expansions (at least in a sector) at some point, for example the origin. There are two general types of approximants (See Nuttall[14] for a detailed treatment). Each of these types involves a set of polynomials. The first type is the so-called Latin polynomial problem, and the second the so-called German polynomial problem. The names apparently trace back to Mahler[15] because roman and gothic type were used for printing the polynomials.

The Latin problem is defined by the equations,

$$\sum_{j=0}^{m} P_j(z) F_j(z) = O(z^{s+1}), \tag{4.1}$$

where, if p_j is the nominal degree of the polynominal $P_j(z)$,

$$s = \sum_{j=0}^{m} (p_j + 1) - 2. \tag{4.2}$$

Approximants can be formed in this way if, for instance, $F_j(z) = [F(z)]^j$ or $F_j(z) = F^{(j)}(z)$ by solving the linear algebraic equation (4.1) for the polynomials and then by solving

$$\sum_{j=0}^{m} P_j(z) y_j(z) = 0, \tag{4.3}$$

under the appropriate boundary conditions, $y(0) = F(0), \ldots$. In these two examples we obtain the algebraic and integral approximants respectively.

The German problem is defined by the equations

$$\mathcal{Q}_j(z)F_i(z) - \mathcal{Q}_i(z)F_j(z) = O(z^{s+3}), \tag{4.4}$$

where the nominal degree of the $\mathcal{Q}_j(z)$ is $s - p_j - 1$. These polynomials (and the Latin ones as well) always exist, but they may not be unique. There are interesting relations between the two sets of polynomials[14] which we will not pursue. If we now choose, for example, $F_0(z) = 1$, and the other $F_j(z)$ as we like, then the set of formulas

$$\mathcal{Q}_j(z) \cdot 1 - \mathcal{Q}_0(z)F_j(z) = O(z^{s+3}), \tag{4.5}$$

is equivalent to simultaneous Padé approximation or vector-valued rational interpolation (See Graves-Morris and Jenkins[16] for a good summary and references, and also Graves-Morris and Saff.[17]) in the form,

$$F_j(z) - \frac{\mathcal{Q}_j(z)}{\mathcal{Q}_0(z)} = O(z^{s+3}), \tag{4.6}$$

as all the approximants have the same denominator. To make a connection with the approximants of section II, we look first at the defining equation (2.7) together with (2.9). We see that the structure (2.2) of the function $f(z)$ is such that we may split (in the neighborhood of $z = \infty$) $f(z) = f_1(z) + f_2(z)$ where each $f_i(z)$ has the corresponding asymptotic behavior of that part of (2.2) with the corresponding subscript. The defining equations (2.7) then become equivalent to (4.6) when we choose

$$F_1(z) = \frac{f_1(z)}{\frac{z^r}{(1+z)^r}\tilde{A}(z)\tilde{B}_1(z)}, \quad F_2(z) = \frac{f_2(z)}{\frac{z^r}{(1+z)^r}\tilde{A}(z)\tilde{B}_2(z)}, \tag{4.7}$$

and the expansion is now made about $z = \infty$ in powers of $1/z$, with an error term $O(z^{-l-1})$. This result shows the relation of quasifractional approximants to the vector-valued, one-point rational interpolant problem. The theory of vector-valued, rational interpolants has been extended to the multi-point problem (hence the name), see for instance Graves-Morris and Saff.[18] The usual extension is to give the vector data for the $F_j(z)$ at additional points. In the case of the quasifractional approximants, eq. (2.6), we use the property that $f(z) = \sum_{i=1}^{2} f_j(z) = \vec{u}\cdot\vec{f}(z)$, where $\vec{u} = \binom{1}{1}$. Thus if we first clear the denominators, the equations (4.6) collapse under this dot product to give,

$$\sum_{i=1}^{2}[\mathcal{Q}_0(z)f_i(z) - \mathcal{Q}_i(z)\frac{z^r}{(1+z)^r}\tilde{A}(z)\tilde{B}_i(z) = O(z^m), \tag{4.8}$$

where an accuracy-through-order condition is now imposed at $z=0$, our choice for a second point (besides ∞). Since the coefficient of the $f_i(z)$'s is independent of i, these recombine into $f(z)$. If we now compare (4.8) with (2.6) we see that in the notation of section II, that (4.8) is just

$$\mathcal{Q}_0(z)f(z) = \mathcal{Q}_0(z)\hat{f}(z) + O(z^m), \tag{4.9}$$

as multiplication by a polynomial can only increase the exponent of z in the order symbol and not decrease it. Thus this condition just completes the defining equations (2.6-9) for the quasifractional approximants when we identify,

$$\mathcal{Q}_1(z) = \sum_{i=0}^{n}\bar{p}_i z^i, \text{ and } \mathcal{Q}_2(z) = \sum_{i=1}^{n}\bar{P}_i z^i. \tag{4.10}$$

If we rewrite (4.8) using $G_0 = f = f_1 + f_2$, $G_1 = -f_1/F_1$ and $G_2 = -f_2/F_2$, we get

$$\sum_{i=0}^{2}\mathcal{Q}_i(z)G_i(z) = O(z^m), \tag{4.11}$$

which differs from (4.1), the Latin polynomial problem only in that the accuracy-through-order principle only partially determines the polynomials (complete determination comes by combination of this equation with that at $z=\infty$), as was the case of the relationship between German polynomial problem and the quasifractional problem at $z=\infty$.

As we have now seen, the structure of the quasifractional approximant problem is basically that of the German-polynomial, Hermite-Padé approximant problem, except that at the origin (where we have a convergent power series) instead of full vector data, only scalar data, derived by the dot product of the vector $\vec{u}$ with the equations is used. We call this process "Latinization" of the problem at one point because as we have seen the equations reduce those for the Latin problem when this procedure is carried out.

In the case where the denominator is predetermined to be $(1+z)^n$, as in (2.5), the corresponding Hermite-Padé problem has the degree of $\mathcal{Q}_0(w)$ as 0 and the degrees of $\mathcal{Q}_1(w)$ and $\mathcal{Q}_2(w)$ as n. The correspondence now comes when we choose $w = z/(1+z)$. If

$$\mathcal{Q}_1(w) = \sum_{i=0}^{n} \check{p}_i w^i, \tag{4.12}$$

then the coefficients in (2.5) are given by

$$p_I = \sum_{J=-\frac{1}{2}I}^{\frac{1}{2}I} \check{p}_{\frac{1}{2}I+J} \binom{n-J-\frac{1}{2}I}{\frac{1}{2}I-J}, \tag{4.13}$$

as can be shown by a a little computation, and similarly for Q_2. The sum here is by integer steps, whether or not the beginning point is integer or half-integer. Thus this sort of predetermined denominator, quasifractional approximant also corresponds to the Hermite-Padé type approximation which we have been discussing. This correspondence between quasifractional and Hermite-Padé approximants makes an extensive body of theory, already cited, available for application to the quasifractional approximation problem.

V. Conclusions

Two-point quasifractional approximants are determined through the power series and the asymptotic expansions at two different points. Those expansions are derived from a given function or from perturbation theory in the case of an eigenvalue problem. This kind of approximant seems more suitable for applications to certain problems in physics than the multi-point Padé method. Our analysis shows how to find them in general cases. These approximants are more specific than the Padé ones in the sense that the independent variable should be specially chosen and the analysis must begin by determining a suitable variable the approximants. However the quasifractional approximants are related to the Hermite-Padé approximants as they represent the one-point Latinization of the German polynomial problem encountered in the theory of vector-valued rational interpolants. We have analyzed this relationship herein.

The structure of our approximant is a combination of several Padé type approximants with auxiliary functions which need to be determined in each case. Our analysis has been illustrated using the particular cases of the Airy functions and Bessel functions of fractional order. The suitable independent variable for the Airy functions is related to the variable in the asymptotic expansions. In the case of the Bessel functions the suitable independent variable is just the usual variable. In both cases the structure of the approximants is given by auxiliary functions, and two Padé type rational approximant forms. For the Airy function on the positive axis, the auxiliary functions are exponential and fractional powers. For the negative real axis for Airy functions, and Bessel functions, the auxiliary functions are trigonometric functions and fractional powers.

A truncation error for these approximants has been found. The simplest form of this error has been found here, and can be applied to some very general cases. Our analysis for the Bessel functions of general fractional order confirms our theory, though the truncation error bound found here is more than ten times larger than the real maximum absolute error.

Acknowledgements. This work was supported in part by the US DOE. One of the authors (P.M.) thanks Los Alamos National Laboratory for supporting his stay there.

Figure Captions

Fig. 1 Maximum errors as a function of ν. The full line and the dashed lines are the truncation errors divided by ten as calculated by the theory developed herein. The full line corresponds to $z_0 > 1$ and the dashed line to $z_0 \leq 1$. The point-dash line corresponds to the maximum error of Ref. 10, obtained by direct calculation of the difference between function and the approximant. For negative ν the errors shown are those for $x^{-\nu} J_\nu(x)$.

References

1. G. A. Baker, Jr. and P. R. Graves-Morris, *Padé Approximants Part I: Basic Theory, Encyclopedia of Mathematics and its Applications, Vol. 13* (Cambridge University Press, London, 1981).
2. G. A. Baker, Jr. and P. R. Graves-Morris, *Padé Approximants Part II: Extensions and Applications, Encyclopedia of Mathematics and its Applications, Vol. 14* (Cambridge University Press, London, 1981).
3. A. Bultheel, *Laurent Series and their Padé Approximations* (Birkhäuser Verlag, Basel-Boston, 1987).
4. R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic Press, London-New York, 1973).
5. F. W. J. Olver, *Asymptotic and Special Functions* (Academic Press, London-New York, 1973).
6. E. Chalbaud and P. Martin, J. Math. Phys. **27**, 699 (1986).
7. P. Martin, Proc. Int. Conf. Plasma Physics, Göteborg, Sweden **1**, 325 (1982).
8. P. Martin and G. Donoso, Phys. Fluids B **1**, 247 (1989).
9. P. Martin and A. L. Guerrero, J. Math. Phys. **26**,705 (1985).
10. P. Martin and A. L Guerrero, J. Comp. Phys. **84** to appear.
11. K. Visentin and P. Martin, J. Math. Phys. **28**, 330 (1987).
12. C. Hermite, Ann. Math., Sér. 2 **21**, 289 (1893).
13. H. Padé (Thesis) Ann. Sci. École Norm. Sup. Suppl. [3] **9**, 1 (1892); — , J. Math., Sér. 4 **10**, 291 (1894).
14. J. Nuttall, J. Approx. Theory **42**, 299 (1984).

15. K. Mahler, Comp. Math. **19**, 95 (1968).

16. P. R. Graves-Morris and C. D. Jenkins, Constr. Approx.**2**, 263 (1986).

17. P. R. Graves-Morris and E. B. Saff, J. Comp. Appl. Math. **23**, 63 (1988).

18. P. R. Graves-Morris and E. B. Saff, in *Rational Approximation and Interpolation, LNIM 1105*, P. R. Graves-Morris, E. B. Saff and R. S. Varga, eds. (Springer, Berlin-New York, 1984) pp. 227-242. æ